

THE DISTRIBUTION OF THE SPINE OF A FLEMING-VIOT TYPE PROCESS

MARIUSZ BIENIEK AND KRZYSZTOF BURDZY

ABSTRACT. We show uniqueness of the spine of a Fleming-Viot particle system under minimal assumptions on the driving process. If the driving process is a continuous time Markov process on a finite space, we show that asymptotically, when the number of particles goes to infinity, the branching rate for the spine is twice that of a generic particle in the system, and every side branch has the distribution of the unconditioned generic branching tree.

1. INTRODUCTION

It is well known that, under suitable assumptions, a branching process can be decomposed into a spine and side branches. A detailed review of the relevant literature is presented in [8, Sect. 2.2]. The “Evans’ immortal particle picture” was introduced in [9]. Another key paper in the area is [11]. Heuristically speaking, the spine has the distribution of the driving process conditioned on non-extinction, the side branches have the distributions of the critical branching process, and the branching rate along the spine is twice the rate along any other trajectory.

We will prove results for the Fleming-Viot branching process introduced in [5] that have the same intuitive content. Our results have to be formulated in a way different from the informal description given above for two reasons. The first, rather mundane, reason is that the Fleming-Viot branching process has a different structure from the processes considered in [8, Sect. 2.2]. A more substantial difference is that for a Fleming-Viot process with a fixed (finite) number of particles, the distribution of the spine does not have an elegant description (as far as we can tell). On the top of that, unlike in the case of superprocesses, the limit of Fleming-Viot processes, when the number of particles goes to infinity, has not been constructed (and might not exist in any interesting sense). Hence, our results will be asymptotic in nature. We will show that the limit of the spine processes, as the number of particles goes to infinity, has the distribution of the driving process conditioned never to hit the boundary. We will also prove that the rate of branching along the spine converges to twice the rate of a generic particle and the

1991 *Mathematics Subject Classification.* 60G17.

Key words and phrases. Fleming-Viot particle system, spine.

Research supported in part by NSF Grant DMS-1206276.

distribution of a side branch converges to the distribution of a branching process with the limiting branching rate.

Our main results on the asymptotic spine distribution are limited to Fleming-Viot processes driven by continuous time Markov processes on finite spaces. We conjecture that analogous results hold for all Fleming-Viot processes (perhaps under mild technical assumptions).

The paper is organized as follows. Section 2 contains basic definitions. It is followed by Section 3 proving existence of the spine under very weak assumptions, thus significantly strengthening a similar result from [10]. Section 4 shows that a historical process, in the spirit of [7], can be represented as a Fleming-Viot process and satisfies an appropriate limit theorem. Section 5 contains the main theorems on the distribution of the spine, its branching rate, and its side branches. Section 6 shows by example that the results on the spine distribution must have asymptotic character because they do not necessarily hold for a process with a fixed number of particles.

2. BASIC DEFINITIONS

Our main theorems will be concerned with Fleming-Viot processes driven by Markov processes on finite state spaces. Nevertheless we need to consider Fleming-Viot processes with an abstract underlying state space because our proofs will be based on “dynamical historical processes” which are Fleming-Viot processes driven by Markov processes with values in function spaces.

Let E be a topological space and let F be a Borel proper subset of E . We will write $F^c = E \setminus F$. Let Y_t , $t \geq 0$, be a continuous time strong Markov process with state space E whose almost all sample paths are right continuous. For $s \geq 0$, let

$$\tau_{F,s} = \inf \{t > s : Y_t \in F^c\},$$

and assume that $\tau_{F,s}$ is a stopping time with respect to the natural filtration of Y for all $s \geq 0$. We assume that F^c is absorbing, i.e., $Y_t = Y_{\tau_{F,s}}$ for all $t \geq \tau_{F,s}$, a.s.

In most papers on the Fleming-Viot process, either Y is a diffusion in an open subset $F \subset \mathbb{R}^d$ or Y is a continuous time Markov process and E is a countable set, so τ_F is a stopping time in those cases. We recall here that the hitting time of a Borel subset of a topological space by a progressively measurable process is a stopping time (see, e.g., Bass [1]).

We will use θ to denote the usual shift operator but in this section and Section 3 we do not assume that the transition probabilities of Y are time homogeneous. We will always make the following assumptions.

- (i) $\mathbb{P}(s < \tau_{F,s} < \infty \mid Y_s = x) = 1$ for all $x \in F$ and $s \geq 0$.
- (ii) For every $x \in F$ and $s \geq 0$, the conditional distribution of $\tau_{F,s}$ given $\{Y_s = x\}$ has no atoms.

Consider an integer $N \geq 2$ and a family $\{U_k^i, 1 \leq i \leq N, k \geq 1\}$ of jointly independent random variables such that U_k^i has the uniform distribution on the set $\{1, \dots, N\} \setminus \{i\}$.

We will use induction to construct a Fleming-Viot type process $\mathbf{X}_t^N = (X_t^1, \dots, X_t^N)$, $t \geq 0$, with values in F^N . Let $\tau_0 = 0$, suppose that $(X_0^{1,1}, \dots, X_0^{1,N}) \in F^N$, and let

$$X_t^{1,1}, \dots, X_t^{1,N}, \quad t \geq 0, \quad (2.1)$$

be independent and have transition probabilities of the process Y . We assume that processes in (2.1) are independent of the family $\{U_k^i, 1 \leq i \leq N, k \geq 1\}$. Let

$$\tau_1 = \inf \{t > 0 : \exists_{1 \leq i \leq N} X_t^{1,i} \in F^c\}.$$

By assumption (ii), no pair of processes can exit F at the same time, so the index i in the above definition is unique, a.s.

For the induction step, assume that the families

$$X_t^{j,1}, \dots, X_t^{j,N}, \quad t \geq 0,$$

and the stopping times τ_j have been defined for $j \leq k$. For each $j \leq k$, denote by i_j the unique index such that $X_{\tau_j}^{j,i_j} \in F^c$. Let

$$X_{\tau_k}^{k+1,m} = X_{\tau_k}^{k,m} \quad \text{for } m \neq i_k,$$

and

$$X_{\tau_k}^{k+1,i_k} = X_{\tau_k}^{k,U_k^{i_k}}.$$

Let the conditional joint distribution of

$$X_t^{k+1,1}, \dots, X_t^{k+1,N}, \quad t \geq \tau_k,$$

given $\{X_t^{k+1,m}, 0 \leq t \leq \tau_k, 1 \leq m \leq N\}$ and $\{U_k^i, 1 \leq i \leq N, k \geq 1\}$, be that of N independent processes with transition probabilities of Y , starting from $X_{\tau_k}^{k+1,m}$, $1 \leq m \leq N$. Let

$$\tau_{k+1} = \inf \left\{ t > \tau_k : \exists_{1 \leq i \leq N} X_t^{k+1,i} \in F^c \right\}.$$

We define $\mathbf{X}_t^N := (X_t^1, \dots, X_t^N)$ by

$$X_t^m = X_t^{k,m}, \quad \text{for } \tau_{k-1} \leq t < \tau_k, \quad k \geq 1, \quad m = 1, 2, \dots, N.$$

Note that the process \mathbf{X}^N is well defined only up to the time

$$\tau_\infty := \lim_{k \rightarrow \infty} \tau_k$$

which will be called the lifetime of \mathbf{X}^N . We do not assume that $\tau_\infty = \infty$, a.s.

We will suppress the dependence on N in some of our notation.

2.1. Dynamical historical processes. The concept of a dynamical historical process (DHP) was introduced in [10, p. 355] under a different name. We chose the name “dynamical historical process” because the concept of DHP is based on an intuitive idea similar to the “historical process” (see [7]). Heuristically speaking, for each $n \in \{1, \dots, N\}$, $\{H_t^n(s), 0 \leq s \leq t\}$ represents the unique path in the branching structure of the Fleming-Viot process which goes from X_t^n to one of the points X_0^1, \dots, X_0^N along the trajectories of X^1, \dots, X^N and does not jump at times τ_k . Note that the process Y may have jumps so a dynamical historical process $\{H_t^n(s), 0 \leq s \leq t\}$ is not necessarily continuous.

Let \mathcal{A} be the family of all sequences of the form $((a_1, b_1), (a_2, b_2), \dots, (a_k, b_k))$, where $a_i \in \{1, \dots, N\}$ and $b_i \in \mathbb{N}$ for all i . For a sequence $\alpha = ((a_1, b_1), (a_2, b_2), \dots, (a_k, b_k))$ we will write $\alpha + (m, n)$ to denote $((a_1, b_1), (a_2, b_2), \dots, (a_k, b_k), (m, n))$. We will define a function $\mathcal{L} : \{1, \dots, N\} \times [0, \tau_\infty) \rightarrow \mathcal{A}$. We interpret $\mathcal{L}(i, s)$ as a label of X_s^i so, by abuse of notation, we will write $\mathcal{L}(X_s^i)$ instead of $\mathcal{L}(i, s)$. We let $\mathcal{L}(X_s^i) = ((i, 0))$ for all $0 \leq s < \tau_1$ and $1 \leq i \leq N$. If $\mathcal{L}(X_s^i) = \alpha$ for $\tau_{k-1} \leq s < \tau_k$, $i \neq i_k$ and $i \neq U_k^{i_k}$ then we let $\mathcal{L}(X_s^i) = \alpha$ for $\tau_k \leq s < \tau_{k+1}$. Suppose that $i = U_k^{i_k}$ and $\mathcal{L}(X_s^i) = \alpha$ for $\tau_{k-1} \leq s < \tau_k$. Then we let $\mathcal{L}(X_s^i) = \alpha + (i, k)$ and $\mathcal{L}(X_s^{i_k}) = \alpha + (i_k, k)$ for $\tau_k \leq s < \tau_{k+1}$.

Later in the paper we will consider a branching process whose individuals are elements of the set $\mathcal{L}(\{1, \dots, N\} \times [0, \tau_\infty))$. A sequence α_2 will be considered an offspring of α_1 if $\alpha_2 = \alpha_1 + (m, n)$ for some m and n .

Suppose that $\mathcal{L}(X_t^n) = ((a_1, b_1), (a_2, b_2), \dots, (a_k, b_k))$ for some $k \geq 1$. From the definition of \mathcal{L} we easily infer that $0 = b_1 < b_2 < \dots < b_k$ and $\tau_{b_k} \leq t$, so that $0 < \tau_{b_1} < \dots < \tau_{b_k} \leq t < \tau_{b_{k+1}}$. For $\tau_{b_m} \leq s < \tau_{b_{m+1}}$ with $1 \leq m < k$ we define $\chi(n, t, s) = a_m$ and $H_t^n(s) = X_s^{a_m}$, and for $\tau_{b_k} \leq s \leq t$ we define $\chi(n, t, s) = a_k$ and $H_t^n(s) = X_s^{a_k}$. Note that $H_t^n(s) = X_s^{\chi(n, t, s)}$ and $\chi(n, t, t) = n$ for all $1 \leq n \leq N$ and $0 \leq s \leq t$.

We will call $\{H_t^n(s), 0 \leq s \leq t\}$ a dynamical historical process (DHP) corresponding to X_t^n . Note that H_t^n is defined only for $1 \leq n \leq N$ and $0 \leq t < \tau_\infty$.

We will say that a branching event occurred along H_t^k on the interval $[s_1, s_2]$, where $0 \leq s_1 \leq s_2 \leq t$, if there exist $s \in [s_1, s_2]$ and $j \neq k$ such that $\chi(j, t, s) = \chi(k, t, s)$ and $\chi(j, t, s_2) \neq \chi(k, t, s_2)$.

3. EXISTENCE AND UNIQUENESS OF THE SPINE

The spine process will be defined below the statement of Theorem 3.1. Roughly speaking, the spine is the unique DHP that extends from time 0 to time τ_∞ . The existence and uniqueness of the spine was proved in [10, Thm. 4] under very restrictive assumptions on the driving process Y and under the assumption that the lifetime τ_∞ is infinite. We will prove that the claim holds under minimal reasonable assumptions, that is, the strong Markov property of the driving process and non-atomic character of the exit time distributions.

Theorem 3.1. *Fix some $N \geq 2$, suppose that Y satisfies assumptions (i)-(ii) in Section 2 and $\mathbf{X}_0^N \in F^N$, a.s. Then, a.s., there exists a unique infinite sequence $((a_1, b_1), (a_2, b_2), \dots)$ such that its every finite initial subsequence is equal to $\mathcal{L}(X_s^i)$ for some $1 \leq i \leq N$ and $s \geq 0$.*

In the notation of the theorem, we define the spine of \mathbf{X}^N by $J(s) = J^N(s) = X_s^{a_m}$ for $\tau_{b_m} \leq s < \tau_{b_{m+1}}$, $m \geq 1$. We also let $\mathcal{L}(J, s) = \mathcal{L}(X_s^{a_m})$ and $\chi(J, s) = a_m$ for $\tau_{b_m} \leq s < \tau_{b_{m+1}}$, $m \geq 1$.

Proof of Theorem 3.1. Step 1. We start with a simple estimate. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a σ -field $\mathcal{G} \subset \mathcal{F}$ and an event $A \in \mathcal{F}$. Suppose that $\mathbb{P}(A) \geq p$ and let $V = \mathbb{P}(A \mid \mathcal{G})$. Then

$$\begin{aligned} p &\leq \mathbb{E}V = \mathbb{E}(V \mathbf{1}_{\{V < p/2\}}) + \mathbb{E}(V \mathbf{1}_{\{V \geq p/2\}}) \leq \frac{p}{2} \mathbb{P}\left(V < \frac{p}{2}\right) + 1 \cdot \mathbb{P}\left(V \geq \frac{p}{2}\right) \\ &\leq \frac{p}{2} + \mathbb{P}\left(V \geq \frac{p}{2}\right). \end{aligned}$$

This implies that

$$\mathbb{P}(\mathbb{P}(A \mid \mathcal{G}) \geq p/2) = \mathbb{P}(V \geq p/2) \geq p/2. \quad (3.1)$$

In the rest of the proof, \mathbb{P} will refer to the probability measure on the probability space where \mathbf{X}^N is defined. Let $\tau_F^i = \inf\{t > 0 : X_t^i \in F^c\}$ and $m_i = \text{median}(\tau_F^i)$ for $1 \leq i \leq N$. Note that since the distribution of each τ_F^i has no atoms, each m_i is uniquely determined. The median m_i depends only on X_0^i . Let \mathbf{j} be the index of the particle with the maximal median of the exit time from F (we choose the smallest of such indices if there is a tie). In other words, \mathbf{j} is the smallest number satisfying $m_{\mathbf{j}} = \max_{1 \leq i \leq N} m_i$.

Recall the definition of τ_k from Section 2. Let i^* be a function of i defined by $\tau_F^{i^*} = \tau_{i^*}$ and let

$$\begin{aligned} A' &= \bigcap_{i \neq \mathbf{j}} \{\tau_F^i \leq m_{\mathbf{j}}, U_{i^*}^i = \mathbf{j}\}, \\ A'' &= \{\tau_F^{\mathbf{j}} > m_{\mathbf{j}}\}, \\ A &= A' \cap A''. \end{aligned}$$

The following estimate holds for any $\mathbf{X}_0^N \in F^N$,

$$\mathbb{P}(A') = \prod_{i \neq \mathbf{j}} [\mathbb{P}(\tau_F^i \leq m_{\mathbf{j}}) \mathbb{P}(U_{i^*}^i = \mathbf{j})] \geq \frac{1}{2^{N-1}} \left(\frac{1}{N-1}\right)^{N-1} =: p > 0.$$

The events A' and A'' are independent and $\mathbb{P}(A'') = 1/2$ so $\mathbb{P}(A) \geq p/2 =: p_1$.

Let $\mathcal{F}_t = \sigma\{\mathbf{X}_s^N, s \leq t\}$. Note that all τ_F^i , $1 \leq i \leq N$, are distinct, a.s., because the hitting time distributions have no atoms. Let $\hat{\tau}^1 < \hat{\tau}^2 < \dots < \hat{\tau}^{N-1}$ be the ordering of the set $\{\tau_F^i, i \neq \mathbf{j}\}$. Let $k(i)$ be defined by $\hat{\tau}^i = \tau_F^{k(i)}$.

Since $\mathbb{P}(A) \geq p_1$, we obtain from (3.1),

$$\mathbb{P}(\mathbb{P}(A \mid \mathcal{F}_{\hat{\tau}^1}) \geq p_1/2) \geq p_1/2.$$

Let $B_1 = \{\mathbb{P}(A \mid \mathcal{F}_{\hat{\tau}^1}) \geq p_1/2\}$ and $C_1 = \{\tau_F^{k(1)} \circ \theta_{\hat{\tau}^1} > m_{\mathbf{j}} - \hat{\tau}^1\}$. If B_1 holds then

$$\mathbf{1}_{\{U_{k(1)^*}^{k(1)} = \mathbf{j}\}} = \mathbb{P}(U_{k(1)^*}^{k(1)} = \mathbf{j} \mid \mathcal{F}_{\hat{\tau}^1}) \geq \mathbb{P}(A' \mid \mathcal{F}_{\hat{\tau}^1}) \geq \mathbb{P}(A \mid \mathcal{F}_{\hat{\tau}^1}) \geq p_1/2 > 0.$$

So if B_1 holds then $\{U_{k(1)^*}^{k(1)} = \mathbf{j}\}$ holds as well. This and the fact that the processes $\{X_t^{k(1)}, t \geq \hat{\tau}^1\}$ and $\{X_t^{\mathbf{j}}, t \geq \hat{\tau}^1\}$ are conditionally i.i.d. given $\mathcal{F}_{\hat{\tau}^1}$ imply that on the event B_1 ,

$$\mathbb{P}(C_1 \cap A \mid \mathcal{F}_{\hat{\tau}^1}) \geq (p_1/2)^2.$$

We have $\mathbb{P}(B_1) \geq p_1/2$, so

$$\mathbb{P}(C_1 \cap A) \geq (p_1/2)^3. \quad (3.2)$$

Next we will apply induction. Let

$$\begin{aligned} p_n &= (p_{n-1}/2)^3, \quad n \geq 2, \\ C_n &= \bigcap_{i=1}^n \{\tau_F^{k(i)} \circ \theta_{\hat{\tau}^i} > m_{\mathbf{j}} - \hat{\tau}^i\}, \quad 2 \leq n \leq N-1, \\ B_n &= \{\mathbb{P}(C_{n-1} \cap A \mid \mathcal{F}_{\hat{\tau}^n}) \geq p_n/2\}, \quad 2 \leq n \leq N-1. \end{aligned}$$

Suppose that for some $1 \leq n \leq N-2$,

$$\mathbb{P}(C_n \cap A) \geq p_{n+1}.$$

Note that the above inequality holds for $n=1$, by (3.2). This induction assumption and (3.1) imply that,

$$\mathbb{P}(B_{n+1}) = \mathbb{P}(\mathbb{P}(C_n \cap A \mid \mathcal{F}_{\hat{\tau}^{n+1}}) \geq p_{n+1}/2) \geq p_{n+1}/2.$$

If B_{n+1} holds then

$$\mathbf{1}_{\{U_{k(n+1)^*}^{k(n+1)} = \mathbf{j}\}} = \mathbb{P}(U_{k(n+1)^*}^{k(n+1)} = \mathbf{j} \mid \mathcal{F}_{\hat{\tau}^{n+1}}) \geq \mathbb{P}(A' \mid \mathcal{F}_{\hat{\tau}^{n+1}}) \geq \mathbb{P}(A \mid \mathcal{F}_{\hat{\tau}^{n+1}}) \geq p_{n+1}/2 > 0.$$

Hence if B_{n+1} holds then $\{U_{k(n+1)^*}^{k(n+1)} = \mathbf{j}\}$ holds as well. This and the fact that the processes $\{X_t^{k(n+1)}, t \geq \hat{\tau}^{n+1}\}$ and $\{X_t^{\mathbf{j}}, t \geq \hat{\tau}^{n+1}\}$ are conditionally i.i.d. given $\mathcal{F}_{\hat{\tau}^{n+1}}$ imply that on the event B_{n+1} ,

$$\mathbb{P}(C_{n+1} \cap A \mid \mathcal{F}_{\hat{\tau}^{n+1}}) \geq (p_{n+1}/2)^2.$$

We have $\mathbb{P}(B_{n+1}) \geq p_{n+1}/2$, so

$$\mathbb{P}(C_{n+1} \cap A) \geq (p_{n+1}/2)^3.$$

This finishes the induction step. We conclude that

$$\mathbb{P}(C_{N-1} \cap A) \geq (p_{N-1}/2)^3 =: q, \quad (3.3)$$

where $q > 0$ depends only on N .

If $C_{N-1} \cap A$ holds then all DHP paths $H_{m_j}^n$, $1 \leq n \leq N$, must pass through $X_{\hat{\tau}^1}^{\mathbf{j}}$, so they all agree with $H_{\hat{\tau}^1}^{\mathbf{j}}$ on the time interval $[0, \hat{\tau}^1]$ in the sense that $\chi(n, m_j, s) = \chi(\mathbf{j}, m_j, s)$ for all $s \in [0, \hat{\tau}^1]$ and $1 \leq n \leq N$.

Step 2. Let $\sigma_0 = 0$, $\mathbf{j}_1 = \mathbf{j}$, $m_{\mathbf{j}_1,1} = m_{\mathbf{j}}$ and $\sigma_1 = m_{\mathbf{j}_1,1} \wedge \tau_F^{\mathbf{j}_1}$.

Suppose that \mathbf{j}_n and σ_n have been defined for $n = 1, \dots, k$. We define $\tau_F^{i,k+1} = \inf \{t > \sigma_k : X_t^i \in F^c\}$ and $m_{i,k+1}$ to be the median of the conditional distribution of $\tau_F^{i,k+1} - \sigma_k$ given σ_k for $1 \leq i \leq N$. Let \mathbf{j}_{k+1} be the smallest number satisfying $m_{\mathbf{j}_{k+1},k+1} = \max_{1 \leq i \leq N} m_{i,k+1}$. Let

$$\sigma_{k+1} = (m_{\mathbf{j}_{k+1},k+1} + \sigma_k) \wedge \tau_F^{\mathbf{j}_{k+1},k+1}.$$

It is easy to see that $\sigma_k < \tau_\infty$, a.s., for all $k \geq 1$. Let C_{N-1}^k and A^k be defined in the way analogous to C_{N-1} and A but relative to \mathbf{j}_k and $m_{\mathbf{j}_k,k}$. Let D_k be defined by the condition $\mathbf{1}_{D_k} \equiv \mathbf{1}_{C_{N-1}^k \cap A^k} \circ \theta_{\sigma_{k-1}}$ for $k \geq 1$.

By (3.3) and the strong Markov property of \mathbf{X}^N applied at time σ_{k-1} , for each $k \geq 2$,

$$\mathbb{P}(D_k^c \mid D_1^c, \dots, D_{k-1}^c) \leq 1 - q.$$

It follows that for every $k \geq 2$,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^k D_i\right) &= 1 - \mathbb{P}(D_1^c) \mathbb{P}(D_2^c \mid D_1^c) \dots \mathbb{P}(D_k^c \mid D_1^c, \dots, D_{k-1}^c) \\ &\geq 1 - (1 - q)^k. \end{aligned}$$

So if $D_* = \bigcup_{k=1}^\infty D_k$, then $\mathbb{P}(D_*) = 1$, i.e., almost surely at least one of the events D_k occurs. For any $m \geq 1$, the same claim applies to the process \mathbf{X}^N after time τ_m , by the strong Markov property, so if $G_m = \{\mathbf{1}_{D_*} \circ \theta_{\tau_m} = 1\}$, then $\mathbb{P}(G_m) = 1$ for all $m \geq 1$, and, therefore,

$$\mathbb{P}\left(\bigcap_{m=1}^\infty G_m\right) = 1. \quad (3.4)$$

Fix any $m \geq 1$. By (3.4), with probability 1, there exists k such that $\mathbf{1}_{D_k} \circ \theta_{\tau_m} = 1$ and we let k denote the smallest integer with this property. Let $\eta_k^{(m)} = \sigma_k \circ \theta_{\tau_m} + \tau_m$, $k \geq 1$. The last remark in Step 1 implies that all DHP paths $H_{\eta_k^{(m)}}^n$, $1 \leq n \leq N$, must agree on the time interval $[0, \tau_m]$, a.s., that is, $\chi(j, \eta_k^{(m)}, s) = \chi(1, \eta_k^{(m)}, s)$ for all $s \in [0, \tau_m]$ and $1 \leq j \leq N$. For any $t < \tau_\infty$ we find a random m such that $t \leq \tau_m$ and k such that $\mathbf{1}_{D_k} \circ \theta_{\tau_m} = 1$. Suppose that $\mathcal{L}\left(X_{\eta_k^{(m)}}^1\right) = ((c_1, d_1), \dots, (c_n, d_n))$. It is easy to check that if we let $a_m = c_m$ and $b_m = d_m$ for $1 \leq m \leq n$ such that $\tau_{d_{m+1}} < t$ then this definition is consistent when we vary t over $[0, \tau_\infty)$. We have defined a unique sequence $((a_1, b_1), (a_2, b_2), \dots)$ satisfying the theorem. \square

4. DYNAMICAL HISTORICAL PROCESS AS A FLEMING-VIOT PROCESS

We will write $\{Y^t, 0 \leq s \leq t\}$ to denote the process Y conditioned by $\tau_{F,0} > t$. In this section, we prove that, as the number of particles N goes to infinity, the empirical distribution of DHPs at time t converges to the distribution of the trajectory of the process Y conditioned by $\tau_{F,0} > t$. For technical reasons we impose two extra assumptions on \mathbf{X} ; they will stay in force for the rest of the paper. We assume that $\tau_\infty = \infty$, a.s., for all N , and that the process Y is time homogeneous. If the driving process is Brownian motion in a Lipschitz domain with the Lipschitz constant less than 1 ([2, 10]) or Brownian motion in a polytope with $N = 2$ ([2]), then $\tau_\infty = \infty$, a.s. However, it was proved in [3] that $\tau_\infty < \infty$, a.s., for every N , for some Fleming-Viot processes driven by one-dimensional diffusions. Crucially for the rest of our paper, it is easy to see that if the driving process is a continuous time Markov process on a finite space then $\tau_\infty = \infty$, a.s.

Let $D([0, t], E)$ denote the usual Skorokhod space of cadlag functions with values in E . Let $\mathbf{H}_t^N = (H_t^1, \dots, H_t^N)$, where H_t^k is DHP of X_t^k . Let \mathcal{X}_t^N and \mathcal{H}_t^N denote empirical distribution of \mathbf{X}_t^N and \mathbf{H}_t^N , resp., i.e.,

$$\begin{aligned}\mathcal{X}_t^N(A) &= \frac{1}{N} \sum_{k=1}^N \delta_{X_t^k}(A), & \text{for } A \subset E, \\ \mathcal{H}_t^N(A) &= \frac{1}{N} \sum_{k=1}^N \delta_{H_t^k}(A), & \text{for } A \subset D([0, t], E).\end{aligned}$$

Let $\mathbb{P}^\mathcal{X}$ and $\mathbb{E}^\mathcal{X}$ denote the probability distribution and the corresponding expectation for the process \mathbf{X}^N , assuming that the empirical distribution of \mathbf{X}_0^N is \mathcal{X} .

Theorem 4.1. *Assume that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$ for some probability measure \mathcal{X} on F and $\tau_\infty = \infty$, a.s., for all N . Then for every fixed $t \geq 0$ and continuous bounded function f on $D([0, t], E)$, when $N \rightarrow \infty$, in probability,*

$$\mathcal{H}_t^N(f) \rightarrow \mathbb{E}^\mathcal{X}(f(\{Y_s^t, 0 \leq s \leq t\})).$$

Proof. The theorem follows rather easily from a result of Villemonais [12] but we have to reformulate the problem to be able to apply that theorem in our setting. Specifically, we have to represent DHP as a time-homogeneous Markov process.

For $y \in D([0, t], E)$ and $t \geq 0$ let

$$y^{(t)}(s) = \begin{cases} y(s), & \text{for } s < t, \\ y(t), & \text{for } s \geq t, \end{cases}$$

and note that $y^{(t)} \in D([0, \infty), E)$. Let $Z_t = (t, Y^{(t)})$. Note that the process Z is a time-homogeneous Markov process with trajectories in the space $D_* := D([0, \infty), \mathbb{R}_+ \times D([0, \infty), E))$. Let

$$F_Z = \{(t, (v_t(s))_{s \geq 0})_{t \geq 0} \in D_* : v_t(s) \in F \text{ for all } s, t \geq 0\}.$$

The set F_Z^c is absorbing for Z in the sense that if $Z_s \in F_Z^c$ and $s < t$ then $Z_t \in F_Z^c$; but it is not true that $Z_t = Z_s$.

Let $\hat{\mathbf{H}}_t^N = (\hat{H}_t^1, \dots, \hat{H}_t^N)$, where $\hat{H}_t^k(s) = (t, H_t^k(s))$ for $s \leq t$ and $\hat{H}_t^k(s) = (t, H_t^k(t))$ for $s > t$. Then $\hat{\mathbf{H}}_t^N$ is a Fleming–Viot process in $\mathbb{R}_+ \times D([0, \infty), E)$ based on Z . A “particle” in this process jumps to the location of another particle when it hits F_Z^c . For $t \geq 0$ we define the empirical distribution $\hat{\mathcal{H}}_t^N$ of $\hat{\mathbf{H}}_t^N$ as

$$\hat{\mathcal{H}}_t^N(A) = \frac{1}{N} \sum_{k=1}^N \delta_{\hat{H}_t^k}(A), \quad \text{for } A \subset \mathbb{R}_+ \times D([0, \infty), E).$$

Note that for $A \subset D([0, \infty), E)$ we have $\hat{\mathcal{H}}_t^N(\mathbb{R}_+ \times A) = \mathcal{H}_t^N(A)$.

By [12, Thm. 1], for every fixed $t \geq 0$ and continuous bounded function g on D_* , when $N \rightarrow \infty$, in probability,

$$\hat{\mathcal{H}}_t^N(g) \rightarrow \mathbb{E}(g(Z_t) \mid Z_s \in F_Z, 0 \leq s \leq t).$$

This is essentially the assertion of the theorem, cloaked in a different formal statement. \square

Remark 4.2. For later reference we state [12, Thm. 1]. This claim may be also considered a corollary of Theorem 4.1. Assume that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$ for some probability measure \mathcal{X} on F . Then for every fixed $t \geq 0$ and continuous bounded function f on E , when $N \rightarrow \infty$, in probability,

$$\mathcal{X}_t^N(f) \rightarrow \mathbb{E}^{\mathcal{X}}(f(Y_t^t)).$$

5. THE ASYMPTOTIC DISTRIBUTION OF THE SPINE

For the remaining part of the paper we assume that Y is a time-homogeneous continuous-time Markov chain with finite state space $E = \{0, 1, \dots, n\}$. We choose $\{1, \dots, n\}$ to play the role of F . We assume that F is a communicating class in the sense that for all $x, y \in F$, there is a positive probability that Y will visit y before hitting 0 if it starts from x . Recall that J_t^N denotes the spine process defined after the statement of Theorem 3.1.

Recall that $\{Y^t, 0 \leq s \leq t\}$ is the process Y conditioned by $\tau_{F,0} > t$. Let Y^∞ denote the process Y conditioned never to leave F . The process Y^∞ can be described as the spatial component of the space-time Doob’s h -process obtained from $\{(t, Y_t), t \geq 0\}$ by conditioning by the parabolic function h which is 0 on F^c and grows to infinity on F . Alternatively, we may define the distribution of Y^∞ as the limit, as $t \rightarrow \infty$, of distributions of Y^t . We will not provide a more formal construction of Y^∞ because it does not pose any technical challenges in our context.

Theorem 5.1. *Consider a probability measure \mathcal{X} on F and suppose that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$. The distribution of J^N converges to the distribution of Y^∞ with the initial distribution \mathcal{X} when $N \rightarrow \infty$.*

Proof. Consider any $t > 0$. By Theorem 4.1, the empirical distribution of the dynamical historical paths of \mathbf{X}_t^N at time t converges to the distribution of $\{Y^t(s), 0 \leq s \leq t\}$, when $N \rightarrow \infty$. Since the set F is finite, this implies that for every fixed $x \in F$, the empirical distribution of DHPs which end at x at time t converges, as $N \rightarrow \infty$, to the distribution of $\{Y^t(s), 0 \leq s \leq t\}$ conditioned by $\{Y^t(t) = x\}$.

Fix any $u > 0$ and a sequence $s_k \rightarrow \infty$. The results of [6] (see especially (1.1) and Section 4) can be used to show that for any $x_k \in F$, the distributions of $\{Y^{s_k}(s), 0 \leq s \leq u\}$ conditioned by $\{Y^{s_k}(s_k) = x_k\}$ converge, as $k \rightarrow \infty$, to the distribution of $\{Y^\infty(s), 0 \leq s \leq u\}$.

Since the state space F is finite, we can use the diagonal method to show that for any sequence N_m going to infinity we can find a subsequence N_m^* of N_m such that for some $p_{x,k} \geq 0$, we have for all x and k ,

$$\lim_{m \rightarrow \infty} \mathbb{P}(J_{s_k}^{N_m^*} = x) = p_{x,k}. \quad (5.1)$$

Given $\mathbf{X}_{s_k}^{N_m^*}$, every DHP which ends at x at time s_k has the same probability of being the initial part of the spine. This observation, (5.1) and the earlier remarks on the convergence of DHPs and convergence of $\{Y^{s_k}(s), 0 \leq s \leq u\}$ conditioned by $\{Y^{s_k}(s_k) = x_k\}$ imply that the distribution of $\{J_s^{N_m^*}, 0 \leq s \leq u\}$ converges, as $m \rightarrow \infty$, to the distribution of $\{Y^\infty(s), 0 \leq s \leq u\}$. Since u is arbitrary and N_m^* is a subsequence of any sequence N_m , the theorem follows. \square

Let q_{xy} denote elements of the transition rate matrix Q for the process Y and let

$$\lambda_t = \sum_{y \in F} \mathbb{P}^{\mathcal{X}}(Y_t^t = y) q_{y0}. \quad (5.2)$$

Let M_t^m be the number of times that the process X^m branched before time t . More formally, in the notation of Section 2, M_t^m is the number of k such that $\tau_k \leq t$ and $U_k^{i_k} = m$. Let M_t^J be the number of times that the process J branched before time t . More precisely, let M_t^J be the number of k such that $\tau_k \leq t$ and either $i_k = \chi(J, \tau_k)$ or $U_k^{i_k} = \chi(J, \tau_k)$.

Proposition 5.2. *Assume that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$ for some probability measure \mathcal{X} on F . For every fixed m , the distribution of M^m converges to the distribution of the Poisson process with variable intensity λ_t as $N \rightarrow \infty$.*

Proof. Every process M^m is a Poisson process with variable random intensity equal to $\sum_{y \in F} \mathcal{X}_t^N(y) q_{y0}$ at time t . Fix any $t > 0$. Definition (5.2) together with finiteness of F imply that it will suffice to prove that, for all $\varepsilon_1, p_1 > 0$,

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\sup_{s \in [0, t]} \sum_{y \in F} |\mathcal{X}_s^N(y) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y)| \geq \varepsilon_1 \right) \leq p_1. \quad (5.3)$$

Since F is finite, it will be enough to prove that, for all $\varepsilon_1, p_1 > 0$ and $y \in F$,

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\sup_{s \in [0, t]} |\mathcal{X}_s^N(y) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y)| \geq \varepsilon_1 \right) \leq p_1.$$

Suppose to the contrary that there exist $p_1, \varepsilon_1 > 0$ and $y \in F$ such that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\sup_{s \in [0, t]} |\mathcal{X}_s^N(y) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y)| \geq \varepsilon_1 \right) \geq p_1.$$

The set F has cardinality n and $\sum_{y \in F} \mathcal{X}_s^N(y) = \sum_{y \in F} \mathbb{P}^{\mathcal{X}}(Y_s^s = y) = 1$ so the above assumption implies that there exist $p_1, \varepsilon_1 > 0$ and $y^* \in F$ such that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\sup_{s \in [0, t]} (\mathcal{X}_s^N(y^*) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y^*)) \geq \varepsilon_1/n \right) \geq p_1. \quad (5.4)$$

It is easy to see that for any $\varepsilon_1, p_1 > 0$ one can find $\delta \in (0, t)$ so small that for every $y \in F$ and $s \geq 0$ the following holds.

- (1) If the number of k such that $X_s^k = y$ is greater than or equal to j then with probability greater than $1 - p_1$, the number of k such that $X_u^k = y$ for all $u \in [s, s + \delta]$ is greater than $j(1 - \varepsilon_1/(3n))$.
- (2) For all $u \in [s, s + \delta]$,

$$|\mathbb{P}^{\mathcal{X}}(Y_s^s = y) - \mathbb{P}^{\mathcal{X}}(Y_u^u = y)| \leq \varepsilon_1/(3n).$$

Let

$$T = \inf \{s \geq 0 : \mathcal{X}_s^N(y^*) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y^*) \geq \varepsilon_1/n\},$$

$$k_1 = \inf \{k \geq 0 : k\delta \geq T\}.$$

By (5.4) and the strong Markov property applied at T ,

$$\limsup_{N \rightarrow \infty} \mathbb{P} (\mathcal{X}_{k_1\delta}^N(y^*) - \mathbb{P}^{\mathcal{X}}(Y_{k_1\delta}^{k_1\delta} = y^*) \geq 2\varepsilon_1/(3n)) \geq p_1(1 - p_1).$$

Note that $k_1\delta < 2t$ and let $m_1 = \lceil 2t/\delta \rceil + 1$. It follows that for some non-random $0 \leq k \leq m_1$,

$$\limsup_{N \rightarrow \infty} \mathbb{P} (\mathcal{X}_{k\delta}^N(y^*) - \mathbb{P}^{\mathcal{X}}(Y_{k\delta}^{k\delta} = y^*) \geq 2\varepsilon_1/(3n)) \geq p_1(1 - p_1)/m_1.$$

This contradicts Remark 4.2 applied at the time $k\delta$. The contradiction completes the proof. \square

Recall the Prokhorov distance between probability measures on the Skorokhod space (see [4, p. 238]). Convergence in the Prokhorov distance is equivalent to the weak convergence of measures.

Corollary 5.3. *Assume that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$ for some probability measure \mathcal{X} on F . Fix any $t_1, t_2 > 0$. Consider a Fleming-Viot process $\{\widehat{\mathbf{X}}_t, t \geq s\}$ with N particles, where $s \in [0, t_1]$. For any $\varepsilon > 0$ there exist $\delta > 0$ and N_1 such that for all $N \geq N_1$, every $1 \leq m \leq N$, $s \in [0, t_1]$ and all initial distributions of $\widehat{\mathbf{X}}$ satisfying $\left| \widehat{\mathcal{X}}_s(y) - \mathbb{P}^{\mathcal{X}}(Y_s^s = y) \right| \leq \delta$ for all $y \in F$, the Prokhorov distance between the distribution of $\{\widehat{M}_t^m, t \in [s, s + t_2]\}$ and the distribution of the Poisson process with intensity λ_t , given by (5.2), on the interval $[s, s + t_2]$ is less than ε .*

Proof. Suppose that the corollary is false. Then there exist $\varepsilon > 0$ and sequences $(\widehat{\mathbf{X}}^k)_{k \geq 1}$, $(s_k)_{k \geq 1}$, $(\delta_k)_{k \geq 1}$ and $(N_k)_{k \geq 1}$, such that we have $s_k \in [0, t_1]$ for all k , $N_k \rightarrow \infty$, $\delta_k \rightarrow 0$, $\left| \widehat{\mathcal{X}}_{s_k}^k(y) - \mathbb{P}^{\mathcal{X}}(Y_{s_k}^{s_k} = y) \right| \leq \delta_k$ and the Prokhorov distance between the distribution of $\{\widehat{M}_t^{k,m}, t \in [s_k, s_k + t_2]\}$ and the distribution of the Poisson process with variable intensity λ_t on the interval $[s_k, s_k + t_2]$ is greater than ε . By compactness, we can find a convergent subsequence of $(s_k)_{k \geq 1}$. By abuse of notation, we will assume that $s_k \rightarrow s_\infty \in [0, t_1]$. This and the continuity of the transition probabilities of Y imply that $\left| \widehat{\mathcal{X}}_{s_k}^k(y) - \mathbb{P}^{\mathcal{X}}(Y_{s_\infty}^{s_\infty} = y) \right| \rightarrow 0$ for all $y \in F$. Let $\mathbf{X}_t^{N_k} := \widehat{\mathbf{X}}_t^{N_k} \circ \theta_{s_k}$. An application of Proposition 5.2 to processes \mathbf{X}^{N_k} shows that the distribution of $\{\widehat{M}_t^{k,m}, t \in [s_k, s_k + t_2]\}$ converges to the distribution of the Poisson process with variable intensity λ_{t+s_∞} on the interval $[0, t_2]$. This contradicts the assumption made at the beginning of the proof. \square

Theorem 5.4. *Assume that \mathcal{X} is a probability measure on F with $\mathcal{X}(x) > 0$ for all $x \in F$. Suppose that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$. The distribution of M^J converges to the distribution of the Poisson process with intensity $2\lambda_t$ when $N \rightarrow \infty$, where λ_t is given by (5.2).*

We have assumed that $\mathcal{X}(x) > 0$ for all $x \in F$ for technical reasons. We expect the theorem to hold without this assumption.

Proof of Theorem 5.4. Fix any $t_1 > 0$ and let $m_1 = \lceil t_1/\delta \rceil + 1$, where δ will be specified later. It would suffice to prove the following assertions.

- (1) For every $\varepsilon > 0$ there exists $\delta_1 > 0$ such that for every $\delta \in (0, \delta_1)$ there exists N_1 such that for $N \geq N_1$ and $m = 1, 2, \dots, m_1$,

$$\mathbb{P}(M_{m\delta}^J - M_{(m-1)\delta}^J = 1 \mid \mathcal{F}_{(m-1)\delta}) \in [(1 - \varepsilon)2\delta\lambda_{(m-1)\delta}, (1 + \varepsilon)2\delta\lambda_{(m-1)\delta}]. \quad (5.5)$$

- (2) There exist c and $\delta_1 > 0$ such that for all $\delta \in (0, \delta_1)$ there exists N_1 such that for $N \geq N_1$ and $m = 1, 2, \dots, m_1$,

$$\mathbb{P}(M_{m\delta}^J - M_{(m-1)\delta}^J > 1 \mid \mathcal{F}_{(m-1)\delta}) \leq c\delta^2. \quad (5.6)$$

Our strategy will be to prove estimates of the type (5.5)-(5.6) but our argument will be a little bit more complicated.

Since F is finite, we have

$$c_1 := n \sup_{x \in F, y \in E} q_{xy} < \infty. \quad (5.7)$$

It is easy to see that for every probability measure \mathcal{X} on F with $\mathcal{X}(x) > 0$ for all $x \in F$ and $\varepsilon > 0$ there exist $c_2, \delta_1 > 0$ and $c_3 < \infty$ such that for all $\delta \in (0, \delta_1]$, $t \in [0, 2t_1]$, and $x \in F$, if Y_0 has the distribution \mathcal{X} then,

$$\mathbb{P}(Y_t^t = x) \geq c_2, \quad (5.8)$$

$$\sup_{s \in [t, t+\delta]} \mathbb{P}(Y_s^s = x) \leq (1 + \varepsilon) \inf_{s \in [t, t+\delta]} \mathbb{P}(Y_s^s = x), \quad (5.9)$$

$$\sup_{s \in [t, t+\delta]} \lambda_s \leq (1 + \varepsilon) \inf_{s \in [t, t+\delta]} \lambda_s, \quad (5.10)$$

$$c_2 \leq \lambda_t \leq c_3. \quad (5.11)$$

Let $M_t = \sum_{m=1}^N M_t^m$. By (5.3) and (5.10), for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}((1 - \varepsilon)N\delta\lambda_{(m-1)\delta} \leq M_{m\delta} - M_{(m-1)\delta} \leq (1 + \varepsilon)N\delta\lambda_{(m-1)\delta}) \geq 1 - \delta^4. \quad (5.12)$$

By (5.3), for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}\left(\sup_{x \in F} \sup_{s \in [(m-1)\delta, m\delta]} \frac{|\mathcal{X}_s^N(x) - \mathbb{P}(Y_s^s = x)|}{\mathbb{P}(Y_s^s = x)} \leq \varepsilon\right) \geq 1 - \delta^4, \quad (5.13)$$

and, trivially following from the last formula,

$$\mathbb{P}\left(\sup_{x \in F} \frac{|\mathcal{X}_{(m-1)\delta}^N(x) - \mathbb{P}(Y_{(m-1)\delta}^{(m-1)\delta} = x)|}{\mathbb{P}(Y_{(m-1)\delta}^{(m-1)\delta} = x)} \leq \varepsilon\right) \geq 1 - \delta^4. \quad (5.14)$$

Let A_1 be the event in the last formula.

Let \tilde{R}_1 be the number of k such that there was exactly one branching event and at least one jump along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$, and $\chi(k, m\delta, m\delta) = \chi(k, m\delta, (m-1)\delta)$. Recall the definition of c_1 from (5.7). The intensity of jumps of any process X^j at any position is bounded by $c_1 < \infty$. It follows that for any $\varepsilon > 0$, the probability that a process X^j will jump at least once on the interval $[(m-1)\delta, m\delta]$ and some other process will jump onto X^j on the same interval (i.e., $U_k^{j_k} = j$ for some $\tau_k \in [(m-1)\delta, m\delta]$) is bounded by $2c_1\delta \cdot 2c_1\delta$, for small δ . We can assume that $\varepsilon < 1$ in (5.12) so we see that there exists c_4 such that for some $\delta_1 > 0$ and all $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{E}\tilde{R}_1 \leq c_4 N \delta^2, \quad \mathbb{E}\tilde{R}_1^2 \leq c_4 N \delta^2.$$

This and (5.11) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$, there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(\tilde{R}_1 \geq N\varepsilon\delta) = \mathbb{P}(\tilde{R}_1^2 \geq (N\varepsilon\delta)^2) \leq \frac{c_4 N \delta^2}{(N\varepsilon\delta)^2} = \frac{c_4}{N\varepsilon^2} \leq \delta^4. \quad (5.15)$$

Let \hat{R}_1 be the number of k such that there was exactly one branching event and at least one jump along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$, and at least one jump occurred later

than the branching event. An argument very similar to that leading to (5.15) shows that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(\widehat{R}_1 \geq N\varepsilon\delta) \leq \delta^4. \quad (5.16)$$

Let R_1 be the number of k such that there was exactly one branching event and at least one jump along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$. Then $R_1 \leq 2\widetilde{R}_1 + \widehat{R}_1$. We combine (5.15) and (5.16) to conclude that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(R_1 \geq 3N\varepsilon\delta) \leq 2\delta^4. \quad (5.17)$$

Since $\varepsilon > 0$ is arbitrarily small in the last formula, (5.8), (5.11), (5.14) and (5.17) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}\left(R_1 \geq N\varepsilon\delta\lambda_{(m-1)\delta} \inf_{x \in F} \mathcal{X}_{(m-1)\delta}^N(x)\right) \leq 3\delta^4. \quad (5.18)$$

Recall the set \mathcal{A} , function \mathcal{L} , and the notions of branching and offspring for elements of \mathcal{A} introduced in Section 2.1. We will consider a branching process \mathbf{B} whose first generation of individuals consists of $\alpha \in \mathcal{A}$ such that $\alpha = \mathcal{L}(X_{(m-1)\delta}^i)$ for some i . The branching process includes all descendants $\beta \in \mathcal{A}$ of the first generation individuals provided $\beta = \mathcal{L}(X_s^j)$ for some j and $(m-1)\delta \leq s \leq m\delta$. An individual has a pair of offspring on $[(m-1)\delta, m\delta]$ with probability not greater than $2c_1\delta$, for small δ . Otherwise it has no offspring. It follows that the expected number of individuals in the j -th generation is bounded by $N(2c_1\delta)^{j-1}$.

Let R_2 be the number of k such that there were exactly two branching events along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$. Let R_3 be the number of k such that there were exactly three branching events along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$. Let R_4 be the number of individuals in the fifth and higher generations of \mathbf{B} . Unlike in the case of R_1 we do not impose any conditions on the number of jumps.

Note that the inequality $\delta \leq \varepsilon/(8c_1^2)$ is equivalent to $N\varepsilon\delta \geq N(2c_1\delta)^2 + N\varepsilon\delta/2$ so either one implies that $\mathbb{P}(R_2 \geq N\varepsilon\delta) \leq \mathbb{P}(R_2 \geq N(2c_1\delta)^2 + N\varepsilon\delta/2)$.

Standard formulas imply that the variance of the number of individuals in the third generation of \mathbf{B} is bounded by $4N(2c_1\delta)^2$ for small δ . This implies that for any $\varepsilon > 0$, some $\delta_1 \in (0, \varepsilon/(8c_1^2))$, for every $\delta \in (0, \delta_1)$, there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(R_2 \geq N\varepsilon\delta) \leq \mathbb{P}(R_2 \geq N(2c_1\delta)^2 + N\varepsilon\delta/2) \leq \frac{4N(2c_1\delta)^2}{(N\varepsilon\delta/2)^2} = \frac{64c_1^2}{N\varepsilon^2} \leq \delta^4. \quad (5.19)$$

The inequality $\delta \leq \sqrt{\varepsilon}/(4c_1^{3/2})$ is equivalent to $N\varepsilon\delta \geq N(2c_1\delta)^3 + N\varepsilon\delta/2$ so either one implies that $\mathbb{P}(R_3 \geq N\varepsilon\delta) \leq \mathbb{P}(R_3 \geq N(2c_1\delta)^3 + N\varepsilon\delta/2)$.

The variance of the number of individuals in the fourth generation of \mathbf{B} is bounded by $4N(2c_1\delta)^3$ for small δ , so for any $\varepsilon > 0$, some $\delta_1 \in (0, \sqrt{\varepsilon}/(4c_1^{3/2}))$, for every $\delta \in (0, \delta_1)$, there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(R_3 \geq N\varepsilon\delta) \leq \mathbb{P}(R_3 \geq N(2c_1\delta)^3 + N\varepsilon\delta/2) \leq \frac{4N(2c_1\delta)^3}{(N\varepsilon\delta/2)^2} = \frac{128c_1^3\delta}{N\varepsilon^2} \leq \delta^4. \quad (5.20)$$

For some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{E}(R_4) \leq \sum_{j \geq 4} N(2c_1\delta)^j \leq 2N(2c_1\delta)^4.$$

This and (5.11) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(R_4 \geq N\varepsilon\delta) \leq \frac{2N(2c_1\delta)^4}{N\varepsilon\delta} = 4c_1^4\delta^3/\varepsilon \leq \delta^{5/2}. \quad (5.21)$$

The same justification which enabled us to conclude (5.18) from (5.17) also gives the following estimates, based on (5.19), (5.20) and (5.21). For any $\varepsilon > 0$ there exists $\delta_1 > 0$ such that for all $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}\left(R_2 \geq N\varepsilon\delta\lambda_{(m-1)\delta} \inf_{x \in F} \mathcal{X}_{(m-1)\delta}^N(x)\right) \leq \delta^4, \quad (5.22)$$

$$\mathbb{P}\left(R_3 \geq N\varepsilon\delta\lambda_{(m-1)\delta} \inf_{x \in F} \mathcal{X}_{(m-1)\delta}^N(x)\right) \leq \delta^4, \quad (5.23)$$

$$\mathbb{P}\left(R_4 \geq N\varepsilon\delta\lambda_{(m-1)\delta} \inf_{x \in F} \mathcal{X}_{(m-1)\delta}^N(x)\right) \leq \delta^{5/2}. \quad (5.24)$$

Let $R_{5,x}$ be the number of particles that jumped to x on the interval $[(m-1)\delta, m\delta]$. Since the particles which exit from F jump to the position of a uniformly chosen particle in F , (5.8), (5.9), (5.12) and (5.13) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}\left(\sup_{x \in F} \frac{|R_{5,x} - \delta\lambda_{(m-1)\delta}N\mathcal{X}_{(m-1)\delta}^N(x)|}{\delta\lambda_{(m-1)\delta}N\mathcal{X}_{(m-1)\delta}^N(x)} \leq \varepsilon\right) \geq 1 - \delta^4. \quad (5.25)$$

Let A_2 be the event in the last formula.

Let $R_{1,x}$ be the number of k such that there was exactly one branching event along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$, and $H_{m\delta}^k(m\delta) = x$. Note that

$$R_{5,x} - R_1 - R_2 - R_3 - R_4 \leq R_{1,x} \leq 2R_{5,x} + R_1.$$

This, (5.18), (5.22), (5.23), (5.24) and (5.25), imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}\left(\sup_{x \in F} \frac{|R_{1,x} - 2\delta\lambda_{(m-1)\delta}N\mathcal{X}_{(m-1)\delta}^N(x)|}{\delta\lambda_{(m-1)\delta}N\mathcal{X}_{(m-1)\delta}^N(x)} \leq 5\varepsilon\right) \geq 1 - 2\delta^{5/2}. \quad (5.26)$$

Let A_3 be the event in the last formula.

Let $R_{6,x}$ be the number of particles that were located at x at time $(m-1)\delta$ and jumped from x to some other state on the interval $[(m-1)\delta, m\delta]$. Recall the definition (5.7) of c_1 and that of the event A_1 in (5.14). We use (5.8) and (5.14) to see that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\begin{aligned} \mathbb{P} \left(\sup_{x \in F} \frac{R_{6,x}}{2c_1 \delta N \mathcal{X}_{(m-1)\delta}^N(x)} \geq 1 \right) &\leq \mathbb{P}(A_1^c) + \mathbb{P} \left(\sup_{x \in F} \frac{R_{6,x}}{2c_1 \delta N \mathcal{X}_{(m-1)\delta}^N(x)} \geq 1 \mid A_1 \right) \\ &\leq \delta^4 + n \max_{x \in F} \mathbb{E} \left(\frac{c_1 \delta N \mathcal{X}_{(m-1)\delta}^N(x)}{(c_1 \delta N \mathcal{X}_{(m-1)\delta}^N(x))^2} \mid A_1 \right) \leq 2\delta^4. \end{aligned} \quad (5.27)$$

Let A_4 be the event on the left hand side in the last formula.

Suppose that $A_2 \cap A_4$ holds. Then for every $x \in F$,

$$N \mathcal{X}_{m\delta}^N(x) \leq N \mathcal{X}_{(m-1)\delta}^N(x) + R_{5,x} \leq N \mathcal{X}_{(m-1)\delta}^N(x) + (1 + \varepsilon) \delta \lambda_{(m-1)\delta} N \mathcal{X}_{(m-1)\delta}^N(x)$$

and

$$N \mathcal{X}_{m\delta}^N(x) \geq N \mathcal{X}_{(m-1)\delta}^N(x) - R_{6,x} \geq N \mathcal{X}_{(m-1)\delta}^N(x) - 2c_1 \delta N \mathcal{X}_{(m-1)\delta}^N(x).$$

Hence,

$$1 - 2c_1 \delta \leq \frac{N \mathcal{X}_{m\delta}^N(x)}{N \mathcal{X}_{(m-1)\delta}^N(x)} \leq 1 + (1 + \varepsilon) \delta \lambda_{(m-1)\delta}.$$

If in addition A_3 holds then

$$\frac{(2 - 5\varepsilon) \delta \lambda_{(m-1)\delta}}{1 + (1 + \varepsilon) \delta \lambda_{(m-1)\delta}} \leq \frac{R_{1,x}}{N \mathcal{X}_{m\delta}^N(x)} \leq \frac{(2 + 5\varepsilon) \delta \lambda_{(m-1)\delta}}{1 - 2c_1 \delta}. \quad (5.28)$$

Let $A_5 = A_2 \cap A_3 \cap A_4$. By (5.25), (5.26) and (5.27), we have $\mathbb{P}(A_5^c) \leq 3\delta^{5/2}$ for small δ , so for every m there exists an event $B_m \in \mathcal{F}_{(m-1)\delta}$ such that

$$\mathbb{P}(B_m) \geq 1 - 2\delta^{5/4} \quad (5.29)$$

and on the event B_m ,

$$\mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \leq 2\delta^{5/4}. \quad (5.30)$$

Let C_k^x be the intersection of $\{H_{m\delta}^k(m\delta) = x\}$ and the event that there was exactly one branching event along $H_{m\delta}^k$ on the interval $[(m-1)\delta, m\delta]$. Let C_J be the event that there was exactly one branching event along the spine on the interval $[(m-1)\delta, m\delta]$. Let $D_k^x = \{\mathcal{L}(J^N, m\delta) = \mathcal{L}(X_{m\delta}^k), J_{m\delta}^N = x\}$. In the following calculation we use the Markov property applied at time $m\delta$ and (5.28),

$$\begin{aligned} \mathbb{P}(C_J \mid \mathcal{F}_{(m-1)\delta}) &= \sum_{x \in F} \sum_{k=1}^N \mathbb{P}(D_k^x \cap C_k^x \mid \mathcal{F}_{(m-1)\delta}) \\ &= \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbb{P}(D_k^x \cap C_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&\quad + \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5^c} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&\leq \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&\leq \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&= \mathbb{E} \left(\sum_{x \in F} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&= \mathbb{E} \left(\sum_{x \in F} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} R_{1,x} \mathbf{1}_{A_5} \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&\leq \mathbb{E} \left(\sum_{x \in F} \mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta}) \frac{(2 + 5\varepsilon)\delta\lambda_{(m-1)\delta}}{1 - 2c_1\delta} \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&= \frac{(2 + 5\varepsilon)\delta\lambda_{(m-1)\delta}}{1 - 2c_1\delta} \mathbb{E} \left(\sum_{x \in F} \mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \\
&= \frac{(2 + 5\varepsilon)\delta\lambda_{(m-1)\delta}}{1 - 2c_1\delta} + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}).
\end{aligned}$$

This, (5.11) and (5.30) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$, on the event B_m ,

$$\begin{aligned}
\mathbb{P}(C_J \mid \mathcal{F}_{(m-1)\delta}) &\leq \frac{(2 + 5\varepsilon)\delta\lambda_{(m-1)\delta}}{1 - 2c_1\delta} + \mathbb{E}(\mathbf{1}_{A_5^c} \mid \mathcal{F}_{(m-1)\delta}) \leq \frac{(2 + 5\varepsilon)\delta\lambda_{(m-1)\delta}}{1 - 2c_1\delta} + 2\delta^{5/4} \\
&\leq (2 + 6\varepsilon)\delta\lambda_{(m-1)\delta}.
\end{aligned} \tag{5.31}$$

Let C'_J be the event that there were at least two branching events along the spine on the interval $[(m-1)\delta, m\delta]$. A calculation similar to that for C_J but using (5.22)-(5.24) instead of (5.26) shows that there exist events $B'_m \in \mathcal{F}_{(m-1)\delta}$ such that $\mathbb{P}(B'_m) \geq 1 - 2\delta^{5/4}$ and for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$, on the event B'_m ,

$$\mathbb{P}(C'_J \mid \mathcal{F}_{(m-1)\delta}) \leq 3\varepsilon\delta\lambda_{(m-1)\delta}. \tag{5.32}$$

We now derive the lower estimate for $\mathbb{P}(C_J \mid \mathcal{F}_{(m-1)\delta})$,

$$\mathbb{P}(C_J \mid \mathcal{F}_{(m-1)\delta}) = \sum_{x \in F} \sum_{k=1}^N \mathbb{P}(D_k^x \cap C_k^x \mid \mathcal{F}_{(m-1)\delta})$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&\quad + \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5^c} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&\geq \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mathbb{P}(D_k^x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \mathbb{E} \left(\sum_{x \in F} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \mathbb{E} \left(\sum_{x \in F} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} \sum_{k=1}^N \mathbf{1}_{C_k^x \cap A_5} \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \mathbb{E} \left(\sum_{x \in F} \frac{\mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta})}{N \mathcal{X}_{m\delta}^N(x)} R_{1,x} \mathbf{1}_{A_5} \mid \mathcal{F}_{(m-1)\delta} \right) \\
&\geq \mathbb{E} \left(\sum_{x \in F} \mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta}) \frac{(2-5\varepsilon)\delta\lambda_{(m-1)\delta}}{1+(1+\varepsilon)\delta\lambda_{(m-1)\delta}} \mathbf{1}_{A_5} \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \frac{(2-5\varepsilon)\delta\lambda_{(m-1)\delta}}{1+(1+\varepsilon)\delta\lambda_{(m-1)\delta}} \mathbb{E} \left(\mathbf{1}_{A_5} \sum_{x \in F} \mathbb{P}(J_{m\delta}^N = x \mid \mathcal{F}_{m\delta}) \mid \mathcal{F}_{(m-1)\delta} \right) \\
&= \frac{(2-5\varepsilon)\delta\lambda_{(m-1)\delta}}{1+(1+\varepsilon)\delta\lambda_{(m-1)\delta}} \mathbb{E} (\mathbf{1}_{A_5} \mid \mathcal{F}_{(m-1)\delta}).
\end{aligned}$$

This, (5.11) and (5.30) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$, on the event B_m ,

$$\begin{aligned}
\mathbb{P}(C_J \mid \mathcal{F}_{(m-1)\delta}) &\geq \frac{(2-5\varepsilon)\delta\lambda_{(m-1)\delta}}{1+(1+\varepsilon)\delta\lambda_{(m-1)\delta}} \mathbb{E} (\mathbf{1}_{A_5} \mid \mathcal{F}_{(m-1)\delta}) \\
&\geq \frac{(2-5\varepsilon)\delta\lambda_{(m-1)\delta}}{1+(1+\varepsilon)\delta\lambda_{(m-1)\delta}} (1-2\delta^{5/4}) \geq (2-6\varepsilon)\delta\lambda_{(m-1)\delta}.
\end{aligned} \tag{5.33}$$

Let $B_* = \bigcap_{1 \leq m \leq m_1} (B_m \cap B'_m)$. We have

$$\mathbb{P}(B_*^c) \leq m_1 4\delta^{5/4} = (\lceil t_1/\delta \rceil + 1) 4\delta^{5/4} \leq c_5 \delta^{1/4}.$$

Let \widetilde{M} be a Poisson process with intensity $2\lambda_t$, independent of M^J . We define \widehat{M}^J by setting $\widehat{M}_t^J - \widehat{M}_{(m-1)\delta}^J = M_t^J - M_{(m-1)\delta}^J$ for $t \in [(m-1)\delta, m\delta]$ on the event $B'_m \cap B_m$ for $m \geq 1$. We let $\widehat{M}_t^J - \widehat{M}_{(m-1)\delta}^J = \widetilde{M}_t - \widetilde{M}_{(m-1)\delta}$ for $t \in [(m-1)\delta, m\delta]$ on $(B'_m \cap B_m)^c$ for $m \geq 1$. It follows from (5.31), (5.32) and (5.33) that \widehat{M}^J satisfies (5.5)-(5.6). Hence, the processes \widehat{M}^J converge to the Poisson process with intensity $2\lambda_t$ as $N \rightarrow \infty$. Since $\mathbb{P}(B_*^c)$ can be made arbitrarily small by choosing small δ , we conclude that the processes M^J converge to the Poisson process with intensity $2\lambda_t$ as $N \rightarrow \infty$. \square

The next theorem is concerned with the distribution of a side branch of the spine of Fleming–Viot process. To this aim we consider two branching processes. The first

process, \mathbf{V} , will be the branching version of Y with the deterministic branching rate λ_t defined in (5.2). Then we will define a branching process \mathbf{Z} representing descendants (along historical paths) of one of the components of \mathbf{X} . The constructions are routine but tedious so we will only sketch them.

Fix any probability measure \mathcal{X} on F . Given $x_1 \in F$ and $t_1 \geq 0$, let $\{\hat{Y}_t, t \geq t_1\}$ have the distribution of the process Y started from x_1 at time t_1 . Let τ_F be the exit time of \hat{Y} from F . Let U be an independent random variable with the distribution given by $\mathbb{P}(U > u) = \exp\left(-\int_{t_1}^u \lambda_t dt\right)$ for $u \geq t_1$. We let $\zeta = \tau_F \wedge U$. If $\tau_F < U$ then we let $\mathbf{a} = 0$. Otherwise $\mathbf{a} = 1$. Let $\mathcal{P}(t_1, x_1)$ denote the distribution of $(\{\hat{Y}_t, t_1 \leq t < \zeta\}, \mathbf{a})$.

Let \mathcal{B} be the family of sequences of the form (i_1, i_2, \dots, i_k) , where $i_1 = 0$ and each i_j is either 0 or 1. If $\beta = (i_1, i_2, \dots, i_k)$ then we will write $\beta + 0 = (i_1, i_2, \dots, i_k, 0)$ and $\beta + 1 = (i_1, i_2, \dots, i_k, 1)$. We will say that $\beta + 0$ and $\beta + 1$ are offspring of β .

Fix any $x_1 \in F$ and $t_1 \geq 0$. There exists a branching process $\mathbf{V} = \mathbf{V}^{t_1, x_1}$ starting from a single individual with the following properties. Individuals V^β in \mathbf{V} are indexed by $\beta \in \mathcal{B}$. Every individual V^β is a process $\{V_t^\beta, s_\beta \leq t < t_\beta\}$ for some $0 \leq s_\beta < t_\beta < \infty$. Let $\mathcal{B}_\mathbf{V}$ denote the random set of all indices of all individuals in \mathbf{V} . We always have $(0) \in \mathcal{B}_\mathbf{V}$. We call V^γ an offspring of V^β if and only if γ is an offspring of β . If $\beta \in \mathcal{B}_\mathbf{V}$ then all ancestors of β are also in $\mathcal{B}_\mathbf{V}$. If $\beta \in \mathcal{B}_\mathbf{V}$ has no offspring in $\mathcal{B}_\mathbf{V}$ then we let $\mathbf{a}_\beta = 0$. Otherwise $\mathbf{a}_\beta = 1$. If γ is an offspring of β then $V_{s_\gamma}^\gamma = V_{t_\beta}^\beta$. If $\beta = (0)$ then the distribution of $(\{V_t^\beta, s_\beta \leq t < t_\beta\}, \mathbf{a}_\beta)$ is $\mathcal{P}(t_1, x_1)$. For any other $\beta \in \mathcal{B}_\mathbf{V}$, the conditional distribution of $(\{V_t^\beta, s_\beta \leq t < t_\beta\}, \mathbf{a}_\beta)$ given the distribution of all ancestors of V^β is that of $\mathcal{P}(s_\beta, V_{s_\beta}^\beta)$. Let $\mathcal{D}(t_1, x_1)$ denote the distribution of \mathbf{V}^{t_1, x_1} .

Remark 5.5. We will now argue that the process \mathbf{V} has a finite lifetime a.s. Let $K(t)$ be the number of individuals at time t . Suppose that the process \mathbf{V} starts at time 0 and its starting distribution is randomized so that the position of the unique individual at time 0 has distribution \mathcal{X} . The branching intensity λ_t has been chosen so that the expected number of individuals is constant in time for this initial distribution, i.e., $\mathbb{E}K(t) = 1$ for all $t \geq 0$. This implies that $K(t)$ cannot grow to infinity (in finite or infinite time) with positive probability. It follows that, with probability 1, for some $c_1 < \infty$, there will be arbitrarily large times t_k with $K(t_k) \leq c_1$. A standard argument based on the strong Markov property shows that \mathbf{V} has to become extinct within one unit of time of one of t_k 's (or earlier), a.s. Since this is an almost sure result, it is easy to see that it implies that $\mathbf{V}^{s, x}$ has a finite lifetime, a.s., for every $s \geq 0$ and $x \in F$.

Suppose that $X_{t_1}^k = x_1$ and let $\alpha_0 = \chi(k, t_1, t_1) \in \mathcal{A}$. Let $\mathcal{A}_\mathbf{Z}$ be the family of all descendants α of α_0 in \mathcal{A} such that $\alpha = \mathcal{L}(X_t^j)$ for some $j \in \{1, \dots, N\}$ and $t \geq t_1$. It is elementary to see that there exists a one to one mapping $\Gamma : \mathcal{A}_\mathbf{Z} \rightarrow \mathcal{B}$ with $\Gamma(\alpha_0) = (0)$, preserving parenthood, i.e., $\Gamma(\alpha_1)$ is a parent of $\Gamma(\alpha_2)$ if and only if α_1 is a parent of α_2 .

We choose such a mapping Γ in an arbitrary way. Let $\mathcal{B}_Z = \Gamma(\mathcal{A}_Z)$. We let $\mathbf{Z}^{k,t_1}(t)$ be a branching process with individuals Z^β for $\beta \in \mathcal{B}_Z$. We call Z^β an offspring of Z^γ if and only if β is an offspring of γ . Every individual Z^β is a process $\{Z_t^\beta, s_\beta \leq t < t_\beta\}$ for some $0 \leq s_\beta < t_\beta < \infty$. If $\beta = \Gamma(\alpha)$ and $\alpha = ((a_1, b_1), (a_2, b_2), \dots, (a_m, b_m))$ then $s_\beta = \inf\{t \geq 0 : \chi(a_m, t, t) = \alpha\}$, $t_\beta = \sup\{t \geq 0 : \chi(a_m, t, t) = \alpha\}$ and $Z_t^\beta = X_t^{a_m}$ for $t \in [s_\beta, t_\beta)$. Note that \mathbf{Z}^{k,t_1} has an infinite lifetime if and only if the spine J^N is a part of this branching process.

Consider any $k \geq 1$ and let u_k be the time of the k -th branching point of the spine J^N . Suppose that $\chi(J, u_k) = j_1$ and note that there is a unique $j_2 \in \{1, \dots, N\}$ such that $j_2 \neq j_1$ and $\chi(j_2, u_k, t) = \chi(j_1, u_k, t)$ for all $t < u_k$. Let $\mathbf{Z}_k = \mathbf{Z}^{j_2, u_k}$.

Theorem 5.6. *Assume that \mathcal{X} is a probability measure on F with $\mathcal{X}(x) > 0$ for all $x \in F$. Suppose that $\mathcal{X}_0^N \Rightarrow \mathcal{X}$ as $N \rightarrow \infty$ and consider the process Y^∞ with the initial distribution \mathcal{X} . Fix any $k \geq 1$ and let $\mu_k(dt)$ be the distribution of the time of the k -th jump of a Poisson process with intensity λ_t , independent of Y^∞ . Then the distribution of \mathbf{Z}_k converges, as $N \rightarrow \infty$, to $\int_0^\infty \sum_{x \in F} \mathcal{D}(t, x) \mathbb{P}(Y_t^\infty = x) \mu_k(dt)$.*

Proof. We will use notation and definitions from the proof of Theorem 5.4. Let

$$A'_1 = \left\{ \sup_{x \in F} \frac{|\mathcal{X}_{m\delta}^N(x) - \mathbb{P}(Y_{m\delta}^{m\delta} = x)|}{\mathbb{P}(Y_{m\delta}^{m\delta} = x)} \leq \varepsilon \right\},$$

and note that this is almost the same as the event A_1 in (5.14) except that $m-1$ is replaced with m . Recall that m_1 is defined to be a function of δ in the proof of Theorem 5.4, and δ is specified later in that proof.

Recall that C_J is the event that there was exactly one branching event along the spine on the interval $[(m-1)\delta, m\delta]$. We obtain from (5.29), (5.31) and (5.33) that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(C_J) \geq (2 - 6\varepsilon)\delta\lambda_{(m-1)\delta}(1 - 2\delta^{5/4}), \quad (5.34)$$

$$\mathbb{P}(C_J) \leq (2 + 6\varepsilon)\delta\lambda_{(m-1)\delta} + 2\delta^{5/4}. \quad (5.35)$$

This, (5.11) and (5.14) imply that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for $N \geq N_1$ and $1 \leq m \leq m_1$,

$$\begin{aligned} \mathbb{P}(A'_1 | C_J) &= \frac{\mathbb{P}(A'_1 \cap C_J)}{\mathbb{P}(C_J)} \geq \frac{\mathbb{P}(C_J) - \mathbb{P}((A'_1)^c)}{\mathbb{P}(C_J)} \\ &\geq \frac{(2 - 6\varepsilon)\delta\lambda_{(m-1)\delta}(1 - 2\delta^{5/4}) - \delta^4}{(2 + 6\varepsilon)\delta\lambda_{(m-1)\delta} + 2\delta^{5/4}} \geq 1 - 7\varepsilon. \end{aligned} \quad (5.36)$$

Fix some $s_1 \geq 0$ and $1 \leq \ell \leq N$. Consider the process \mathbf{Z}^{ℓ, s_1} conditioned on $\{X_{s_1}^\ell = Z_{s_1}^{(0)} = x_1\}$. It follows from Proposition 5.2 that the distribution of the first individual in this process, i.e., $\{Z_t^{(0)}, t \geq s_1\}$ converges, as $N \rightarrow \infty$, to the distribution of the process Y killed at the first jump of an independent Poisson process with intensity λ_t , for $t \geq s_1$.

In other words, it converges to the distribution of the first individual in the the process \mathbf{V}^{s_1, x_1} .

Next consider any family of sequences $\{\beta_1 = (0), \beta_2, \dots, \beta_m\} \subset \mathcal{B}$ such that if β_j belongs to the family then the parent belongs to the family as well. By the Markov property applied at the death times of individuals, for any such family, for every β_j , Proposition 5.2 implies that the distribution of the individual labeled β_j in \mathbf{Z}^{ℓ, s_1} converges to the distribution of the individual with the same label in \mathbf{V}^{s_1, x_1} . Moreover, we have convergence of the joint distribution of all individuals labeled $\beta_1, \beta_2, \dots, \beta_m$ in the process \mathbf{Z}^{ℓ, s_1} to the joint distribution of the similarly labeled individuals in \mathbf{V}^{s_1, x_1} . By Remark 5.5, the process \mathbf{V}^{s_1, x_1} has a finite lifetime so $\mathcal{B}_{\mathbf{V}}$ is finite, a.s. This completes the proof that the distribution of \mathbf{Z}^{ℓ, s_1} converges to that of \mathbf{V}^{s_1, x_1} .

By Theorem 5.1 and (5.14), for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$, there exists N_1 so large that for all $N \geq N_1$, $1 \leq \ell \leq N$ and $1 \leq m \leq m_1$,

$$\mathbb{P}(\chi(J, m\delta) = \ell) < \varepsilon. \quad (5.37)$$

We obtain from the above argument concerning the distribution of \mathbf{Z}^{ℓ, s_1} combined with the Markov property applied at $m\delta$, Corollary 5.3, (5.36) and (5.37) that for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$ there exists N_1 so large that for all $x \in F$, $N \geq N_1$, $1 \leq \ell \leq N$ and $1 \leq m \leq m_1$, conditional on C_J and the event $\{Z_{m\delta}^\ell = x, \chi(J, m\delta) \neq \ell\}$, the Prokhorov distance between the distribution of $\mathbf{Z}^{\ell, m\delta}$ and that of $\mathbf{V}^{m\delta, x}$ is less than ε .

Since the definition of j_2 (an element of the definition of \mathbf{Z}_k) does not refer to the post- u_k process, the claim made in the last paragraph applies not only to a fixed ℓ but also to j_2 . Hence, for any $\varepsilon > 0$, some $\delta_1 > 0$, for every $\delta \in (0, \delta_1)$, there exists N_1 so large that for all $N \geq N_1$ and $1 \leq m \leq m_1$, conditional on C_J , the Prokhorov distance between the distribution of \mathbf{Z}_k and that of $\mathbf{V}^{m\delta, x}$ is less than ε . This easily implies the theorem. \square

Remark 5.7. It is not hard to see that the following “propagation of chaos” assertion holds: for any fixed k , the processes $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k$ are asymptotically independent, when $N \rightarrow \infty$.

6. SPINE DISTRIBUTION FOR A FIXED N

The main result of this paper states that if the number N of particles of a Fleming-Viot process increases to infinity then the distribution of the spine converges to the distribution of the underlying Markov process conditioned not to hit the boundary. One could wonder whether the theorem must have the asymptotic character; perhaps the claim is true for every fixed N . This section is devoted to an example of a Fleming-Viot process with $N = 2$ particles such that the distribution of the spine is not the same as the distribution of the driving process conditioned on non-extinction.

Let Y_t be a continuous-time Markov process Y_t with the state space $E = \{0, 1, 2\}$, $F = \{1, 2\}$ and the transition rate matrix

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 4 & -6 & 2 \\ 1 & 6 & -7 \end{bmatrix}.$$

Let $\mathbf{X}_t = (X_t^1, X_t^2)$ denote the 2-particle Fleming-Viot process based on Y . Then \mathbf{X} has the state space $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$ and the transition rate matrix

$$\mathbf{A} = \begin{bmatrix} -4 & 2 & 2 & 0 \\ 7 & -13 & 0 & 6 \\ 7 & 0 & -13 & 6 \\ 0 & 6 & 6 & -12 \end{bmatrix}.$$

Therefore the stationary distribution π of \mathbf{X} determined by $\pi \mathbf{A} = 0$ is

$$\pi = \left(\frac{7}{13}, \frac{2}{13}, \frac{2}{13}, \frac{2}{13} \right). \quad (6.1)$$

Let Y' and Y'' be independent copies of Y . The state space of (Y', Y'') is

$$\Lambda = \{(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), (1, 1), (1, 2), (2, 1), (2, 2)\}$$

and the transition rate matrix for (Y', Y'') is

$$\mathbf{B} = \begin{bmatrix} 0 & & \dots & & & & & & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & & \dots & & & & & & 0 \\ 0 & 4 & 0 & 4 & 0 & -12 & 2 & 2 & 0 \\ 0 & 0 & 4 & 1 & 0 & 6 & -13 & 0 & 2 \\ 0 & 1 & 0 & 0 & 4 & 6 & 0 & -13 & 2 \\ 0 & 0 & 1 & 0 & 1 & 0 & 6 & 6 & -14 \end{bmatrix}.$$

Let

$$f(x, y) = \mathbb{P}(Y'' \text{ reaches } 0 \text{ before } Y' \mid Y'_0 = x, Y''_0 = y), \quad x, y = 1, 2.$$

Then f is harmonic with respect to \mathbf{B} , i.e., for all $(x_1, x_2) \in \Lambda$,

$$\sum_{(y_1, y_2) \in \Lambda} \mathbf{B}((x_1, x_2), (y_1, y_2)) (f(y_1, y_2) - f(x_1, x_2)) = 0.$$

It follows from the definition of f that $f(0, 1) = f(0, 2) = 0$ and $f(1, 0) = f(2, 0) = 1$. By symmetry, $f(1, 1) = \frac{1}{2}$ and $f(2, 2) = \frac{1}{2}$. It is elementary to check that

$$f(1, 2) = \frac{5}{13}, \quad f(2, 1) = \frac{8}{13}.$$

Let J_t denote the spine of \mathbf{X}_t . The spine passes through X_t^1 (i.e., $\chi(J, t) = 1$) if and only if X^2 “jumps to 0” before X^1 , after time t . The probability of this event is the same as the probability that Y'' will hit 0 before Y' , assuming that $(Y'_0, Y''_0) = (X_t^1, X_t^2)$. If

$\mathbf{X}_t = (1, 1)$ then $J_t = 1$, and if $\mathbf{X}_t = (2, 2)$ then $J_t = 2$. If $\mathbf{X}_t = (1, 2)$ then $J_t = 1$ with probability $f(1, 2) = \frac{5}{13}$, and, by symmetry, if $\mathbf{X}_t = (2, 1)$ then $J_t = 1$ with probability $\frac{5}{13}$. Assume that \mathbf{X} is in the stationary regime and recall the stationary probabilities for \mathbf{X} given in (6.1) to see that

$$\mathbb{P}(J_t = 1) = \frac{7}{13} \cdot 1 + \frac{2}{13} \cdot \frac{5}{13} + \frac{2}{13} \cdot \frac{5}{13} + \frac{2}{13} \cdot 0 = \frac{111}{169}. \quad (6.2)$$

We will show that for a (generic) fixed $t > 0$, the distribution of J_t is not the same as the distribution of Y conditioned to stay in F until time t , and it is not the same as the distribution of Y conditioned to stay in F forever.

Let \mathbb{P}_μ denote the distribution of Y with the initial distribution μ and assume that $\mu(0) = 0$. The transition probabilities of Y are given by

$$P_t^Y := e^{tA} = \begin{bmatrix} 1 & 0 & 0 \\ 1 - \frac{6}{7}e^{-3t} - \frac{1}{7}e^{-10t} & \frac{4}{7}e^{-3t} + \frac{3}{7}e^{-10t} & \frac{2}{7}e^{-3t} - \frac{2}{7}e^{-10t} \\ 1 - \frac{9}{7}e^{-3t} + \frac{2}{7}e^{-10t} & \frac{6}{7}e^{-3t} - \frac{6}{7}e^{-10t} & \frac{3}{7}e^{-3t} + \frac{4}{7}e^{-10t} \end{bmatrix},$$

so

$$\mathbb{P}_\mu(Y_t \neq 0) = \left(\frac{6}{7}\mu(1) + \frac{9}{7}\mu(2) \right) e^{-3t} + \left(\frac{1}{7}\mu(1) - \frac{2}{7}\mu(2) \right) e^{-10t}, \quad (6.3)$$

$$\mathbb{P}_\mu(Y_t = 1) = \left(\frac{4}{7}\mu(1) + \frac{6}{7}\mu(2) \right) e^{-3t} + \left(\frac{3}{7}\mu(1) - \frac{6}{7}\mu(2) \right) e^{-10t}, \quad (6.4)$$

$$\mathbb{P}_\mu(Y_t = 2) = \left(\frac{2}{7}\mu(1) + \frac{3}{7}\mu(2) \right) e^{-3t} + \left(\frac{4}{7}\mu(2) - \frac{2}{7}\mu(1) \right) e^{-10t}. \quad (6.5)$$

It follows that

$$\mathbb{P}_\mu(Y_t = 1 \mid Y_t \neq 0) = \frac{\mathbb{P}_\mu(Y_t = 1)}{\mathbb{P}_\mu(Y_t \neq 0)} = \frac{2(2\mu(1) + 3\mu(2)) + (3\mu(1) - 6\mu(2))e^{-7t}}{3(2\mu(1) + 3\mu(2)) + (\mu(1) - 2\mu(2))e^{-7t}} \rightarrow \frac{2}{3}$$

as $t \rightarrow \infty$, regardless of the initial distribution μ . Comparing this value to (6.2), we see that for large t the distribution of J_t (in the stationary regime) is not the same as the law of Y_t conditioned to stay in F until t .

Next we will compare the distribution of J_t with the distribution of the process Y conditioned to stay in F forever, i.e., Y_t^∞ . Since 0 is an absorbing state,

$$\mathbb{P}_\mu(Y_t^\infty = x) = \lim_{s \rightarrow \infty} \mathbb{P}_\mu(Y_t = x \mid Y_{t+s} \neq 0), \quad x = 1, 2, \quad t > 0.$$

By the Markov property,

$$\begin{aligned} \mathbb{P}_\mu(Y_t = 1 \mid Y_{t+s} \neq 0) &= \frac{\mathbb{P}_\mu(Y_t = 1)(1 - P_s^Y(1, 0))}{\mathbb{P}_\mu(Y_t = 1)(1 - P_s^Y(1, 0)) + \mathbb{P}_\mu(Y_t = 2)(1 - P_s^Y(2, 0))} \\ &= \frac{\mathbb{P}_\mu(Y_t = 1)}{\mathbb{P}_\mu(Y_t = 1) + \mathbb{P}_\mu(Y_t = 2) \frac{1 - P_s^Y(2, 0)}{1 - P_s^Y(1, 0)}}. \end{aligned} \quad (6.6)$$

By (6.3),

$$\lim_{s \rightarrow \infty} \frac{1 - P_s^Y(2, 0)}{1 - P_s^Y(1, 0)} = \lim_{s \rightarrow \infty} \frac{9 - 2e^{-7s}}{6 + e^{-7s}} = \frac{3}{2},$$

so this and (6.4)-(6.5) imply that

$$\mathbb{P}_\mu(Y_t^\infty = 1) = \frac{4}{7} + \frac{6}{7} \cdot \frac{\mu(1) - 2\mu(2)}{2\mu(1) + 3\mu(2)} e^{-7t}.$$

This probability converges to $4/7$ when $t \rightarrow \infty$. This value is different from that in (6.2) so for large t the distribution of J_t (in the stationary regime) is not the same as the law of Y_t^∞ .

7. ACKNOWLEDGMENTS

We are grateful to Theodore Cox, Simon Harris, Doug Rizzolo and Anton Wakolbinger for the most helpful advice.

REFERENCES

- [1] R. F. Bass. The measurability of hitting times. *Electron. Commun. Probab.*, 15:99–105, 2010.
- [2] M. Bieniek, K. Burdzy, and S. Finch. Non-extinction of a Fleming-Viot particle model. *Probab. Theory Related Fields*, 153(1-2):293–332, 2012.
- [3] M. Bieniek, K. Burdzy, and S. Pal. Extinction of Fleming-Viot-type particle systems with strong drift. *Electron. J. Probab.*, 17:no. 11, 15, 2012.
- [4] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [5] K. Burdzy, R. Holyst, and P. March. A Fleming-Viot particle representation of the Dirichlet Laplacian. *Comm. Math. Phys.*, 214(3):679–703, 2000.
- [6] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probability*, 4:192–196, 1967.
- [7] D. A. Dawson and E. A. Perkins. Historical processes. *Mem. Amer. Math. Soc.*, 93(454):iv+179, 1991.
- [8] J. Engländer and A. E. Kyprianou. Local extinction versus local exponential growth for spatial branching processes. *Ann. Probab.*, 32(1A):78–99, 2004.
- [9] S. N. Evans. Two representations of a conditioned superprocess. *Proc. Roy. Soc. Edinburgh Sect. A*, 123(5):959–971, 1993.
- [10] I. Grigorescu and M. Kang. Immortal particle for a catalytic branching process. *Probab. Theory Related Fields*, 153(1-2):333–361, 2012.
- [11] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.
- [12] D. Villemonais. General approximation method for the distribution of markov processes conditioned not to be killed. *ESAIM: Probability and Statistics*, 18:441–467, 2014.

MB: INTITUTE OF MATHEMATICS, UNIVERSITY OF MARIA CURIE SKŁODOWSKA, 20-031 LUBLIN, POLAND

KB: DEPARTMENT OF MATHEMATICS, BOX 354350, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195, USA

E-mail address: mariusz.bieniek@umcs.lublin.pl

E-mail address: burdzy@math.washington.edu